

陆地棉FatB蛋白结构分析和关键 活性位点预测

关红辉、吉夏洁、李飞、严玫

博士1班 CAAS group 4

目录

一. 研究背景

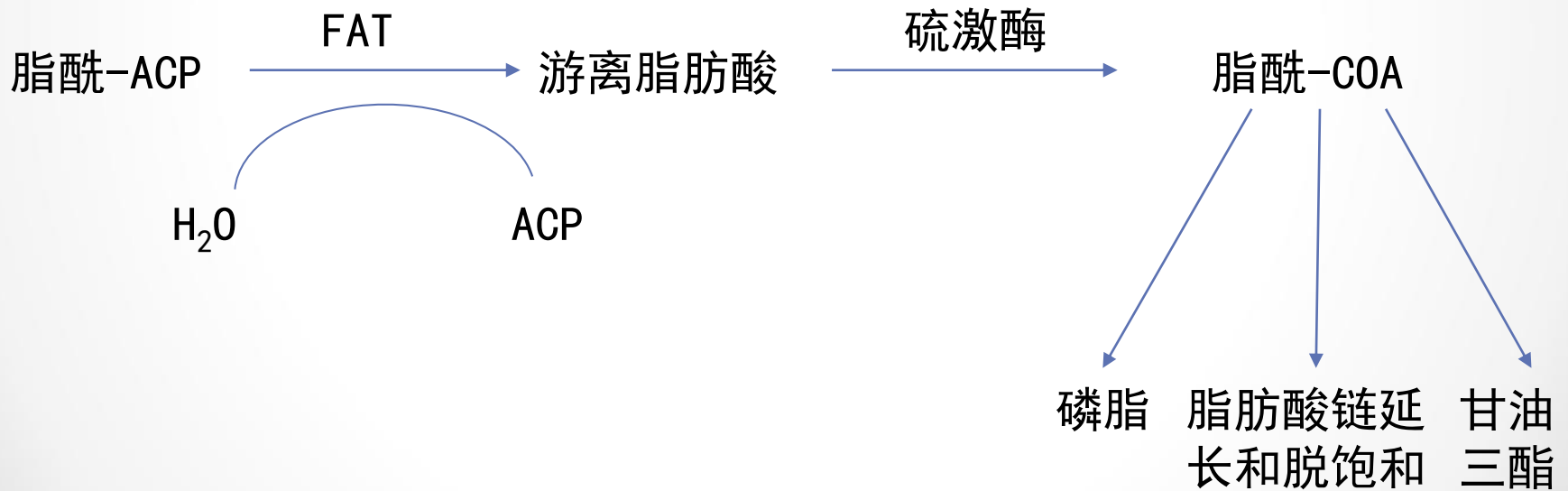
二. 研究方法

三. 结果与分析

四. 结论

一、研究背景

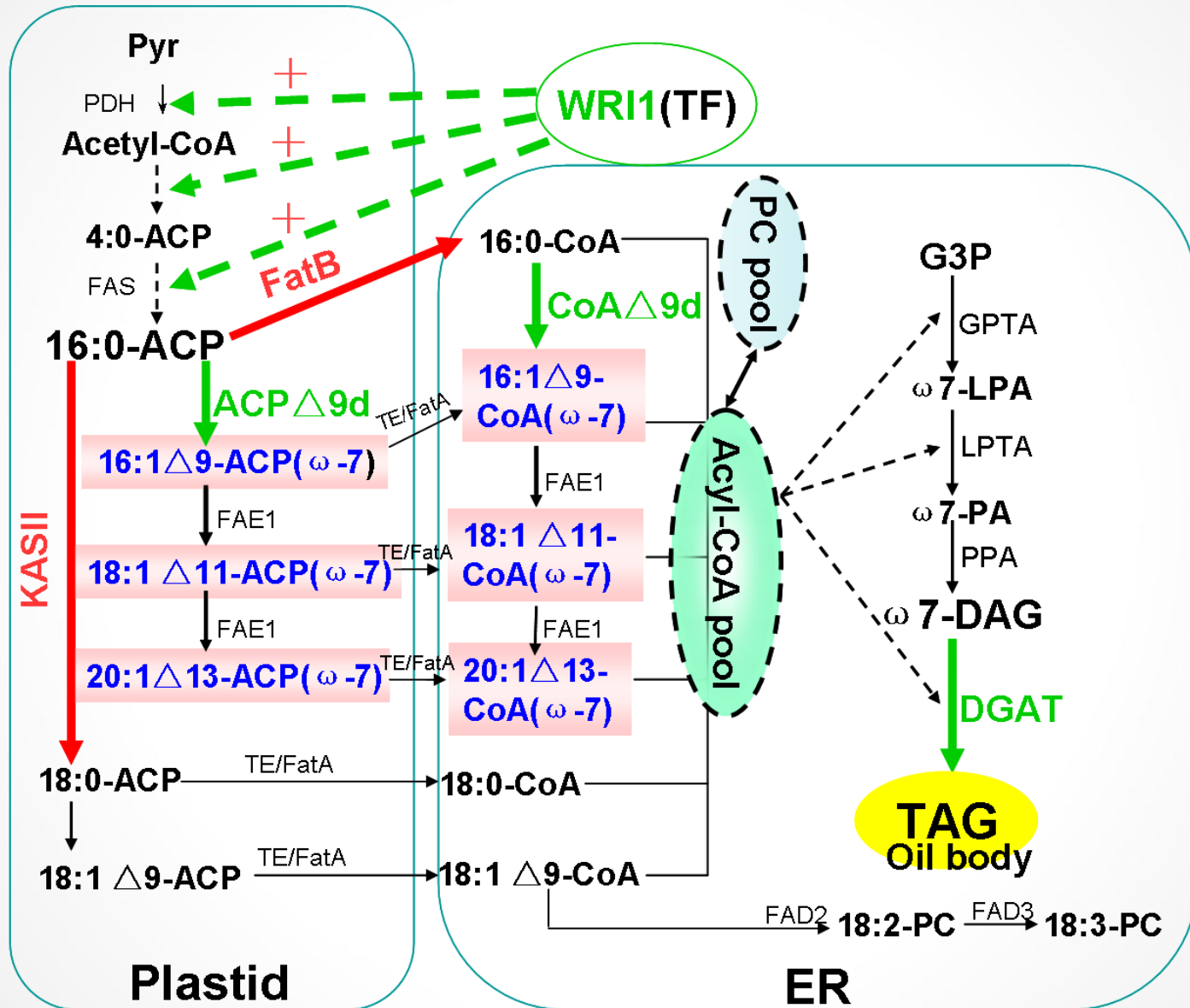
植物细胞内的酰基-ACP硫酯酶是一类质体球蛋白，决定着植物细胞中所合成游离脂肪酸的碳链长度与种类。



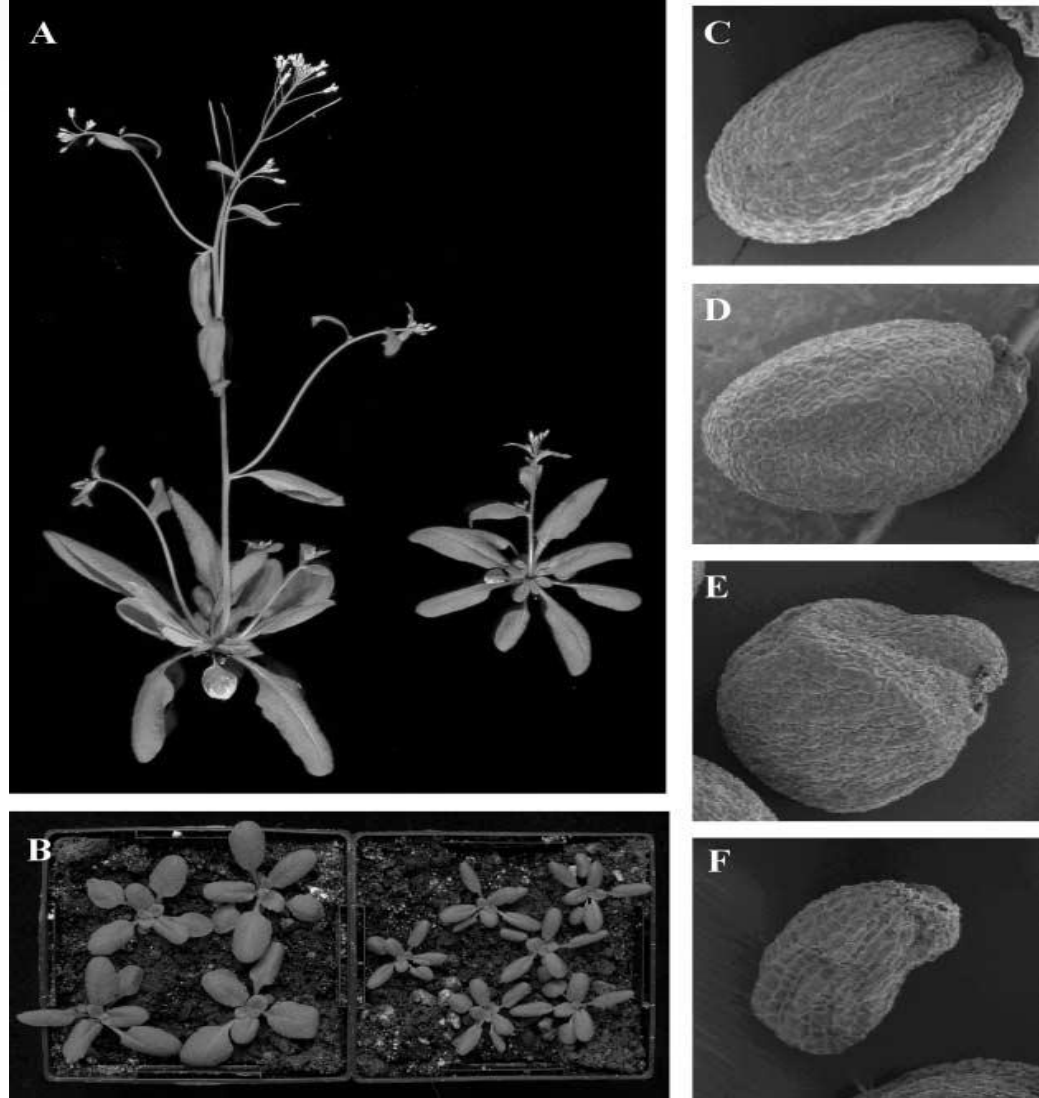
FAT酶催化反应的过程及游离脂肪酸的去向

高等植物的Fat酶主要分为FatA和FatB两类。在大多数植物中，FatA以较高的活性催化18:1-ACP水解。FatB酶主要催化14-18C的饱和脂酰-ACP的水解，尤其对16:0-ACP具有较高活性。

这种底物的特异性直接导致了植物油脂储存在组成和数量上的差异。许多研究利用的这个特性,人为改变其底物特异性,使植物种子可以积累更多特殊的脂肪酸



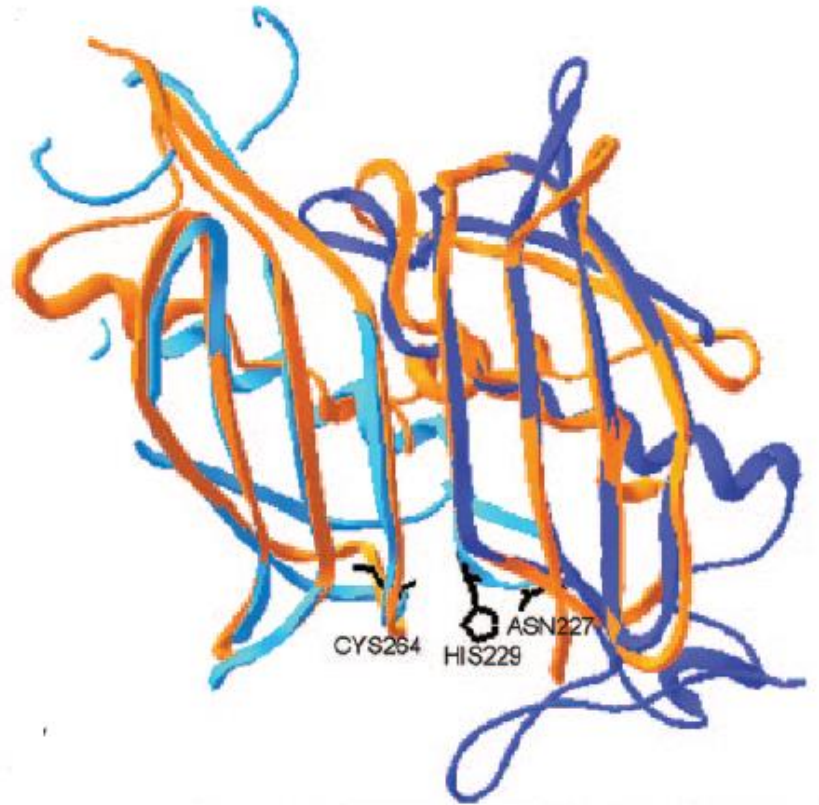
- FatB基因敲除突变体的生长速率降低，植株矮小
- 种子形状和活力也受到影响
- C16:0和C18:0含量减少



(Bonaventure, 2003)

饱和脂肪酸对拟南芥生长发育有很重要的作用

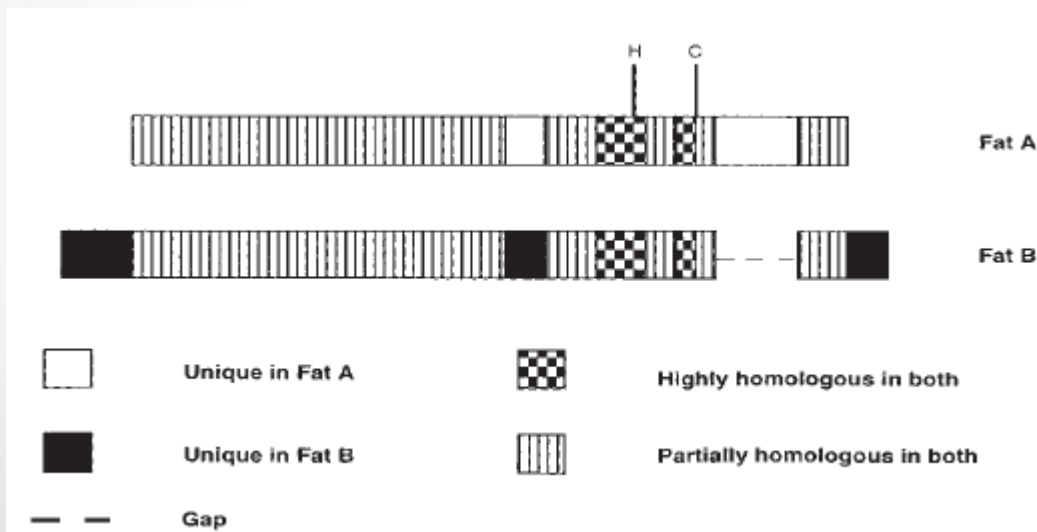
- Mayer (2004) 等成功预测出了AtFatB的三级结构
- Cys-264和His-229的催化残基定位于模型的另一区域，与其他具有螺旋/多股片状结构的酶(热狗折叠)中的催化残基相同。
- 在模型的基础上，ASN-227可能是含有与木瓜蛋白酶类似的催化中心的三位一体的第三个成员。



(Mayer等,2004)

- Mayer 等(2007)预测AtFatB蛋白的四个位置(74、86、141和174)的替换在不同程度上改变了底物特异性。
- 而在110和221两个位置的变化基本上使硫酯酶失活。

- FatB家族的N-末端大约前60个氨基酸代表间质转运肽，它们在两类TES之间和内部几乎没有同源性。在转运肽后，所有的FatB都有一个同源的疏水区，大约有35个氨基酸。
- FatB家族具有一个独特的可变长度的C端片段，该模块的缺失不影响底物特异性，但降低酶活性。这表明，该区域对底物选择性不是关键，但可能在稳定整体结构方面发挥作用。



(Facciotti等, 1998)

目的

- 预测棉花FatB蛋白质的结构和功能。
- 鉴定棉花FatB蛋白中决定其功能的关键氨基酸，为进一步研究其功能提供参考。

二、研究方法

数据库：Uniprot、NCBI

生信分析工具：

- 蛋白质序列比对：

<http://www.ibi.vu.nl/programs/pralinewww/>

- 疏水性：<http://web.expasy.org/protscale/>

- 跨膜结构：<http://www.cbs.dtu.dk/services/TMHMM/>

- 信号肽：<http://www.cbs.dtu.dk/services/SignalP/>

- 结构域：

<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>

- 三级结构预测：<https://www.swissmodel.expasy.org/>

软件：

蛋白质结构分析：SPDBV；系统发育分析：MEGA 7.0

三、结果与分析

序列获得

- 在Uniprot中查找获得蛋白质序列

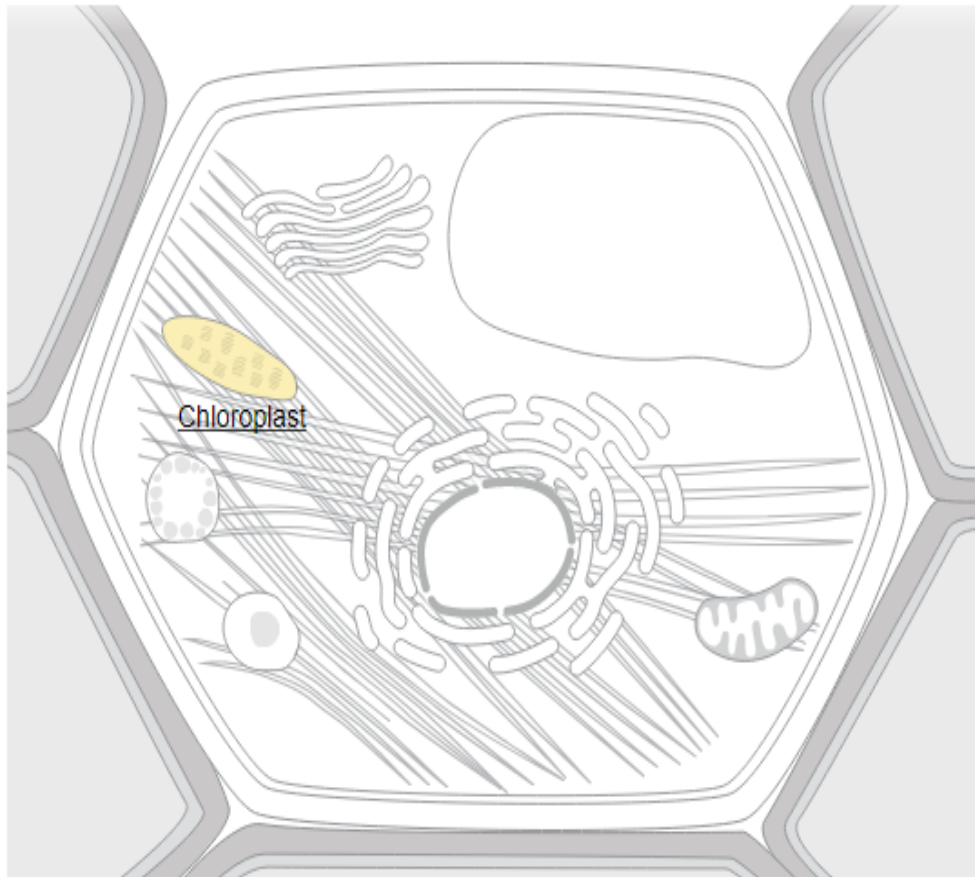
拟南芥FatB (Q9SJE2)

棉花的FatB (Q9SQI3)

拟南芥Fat A1 (Q42561)

拟南芥Fat A2 (Q9SV64)

Subcellular locationⁱ



Graphics by Christian Stolte; Source: COMPARTMENTS

Manual annotation Automatic computational assertion

UniProt annotation

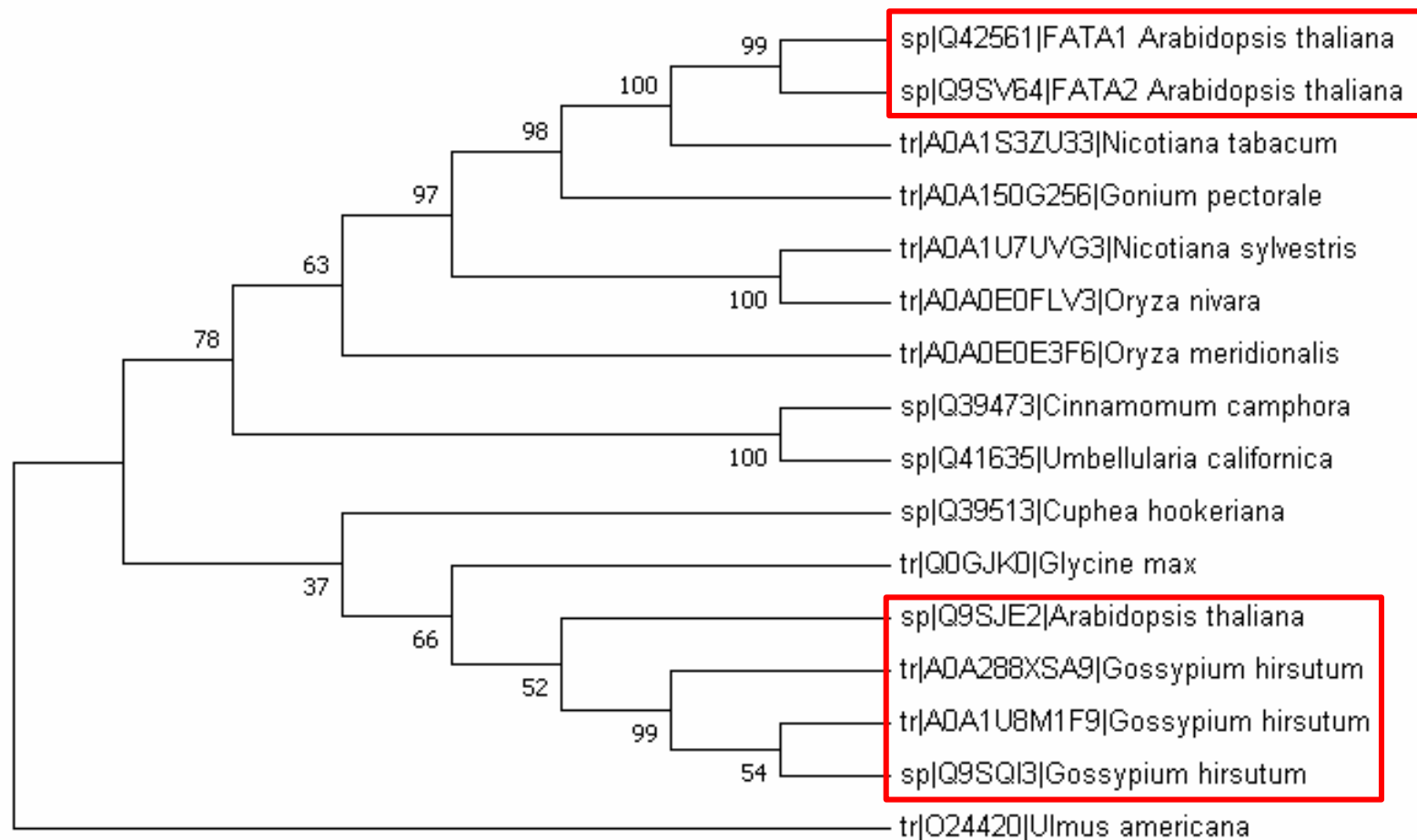
GO - Cellular component

Chloroplast

chloroplast Source: UniProtKB-SubCell

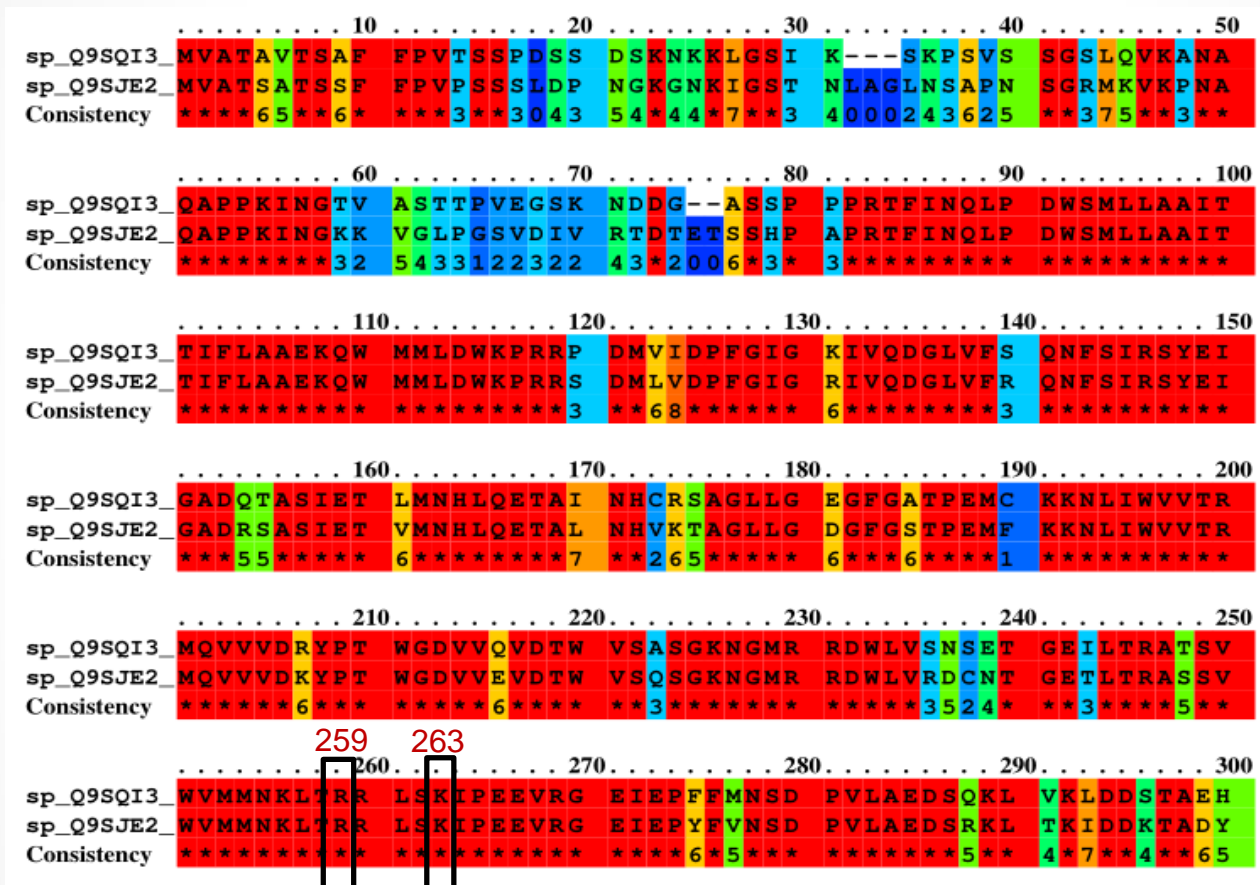
[View the complete GO annotation on QuickGO ...](#)

系统发育分析

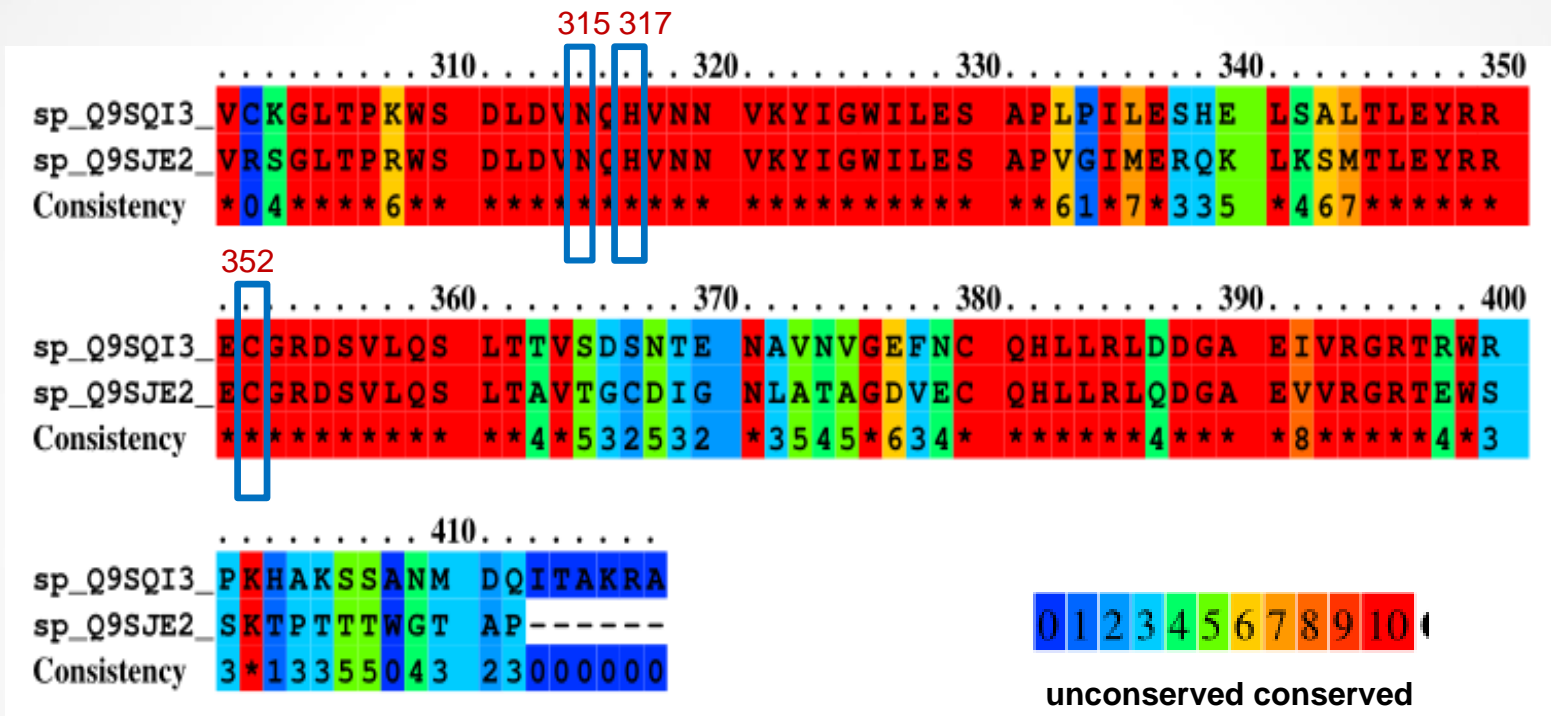


棉花FatB关键保守氨基酸分析

与拟南芥FatB序列比对

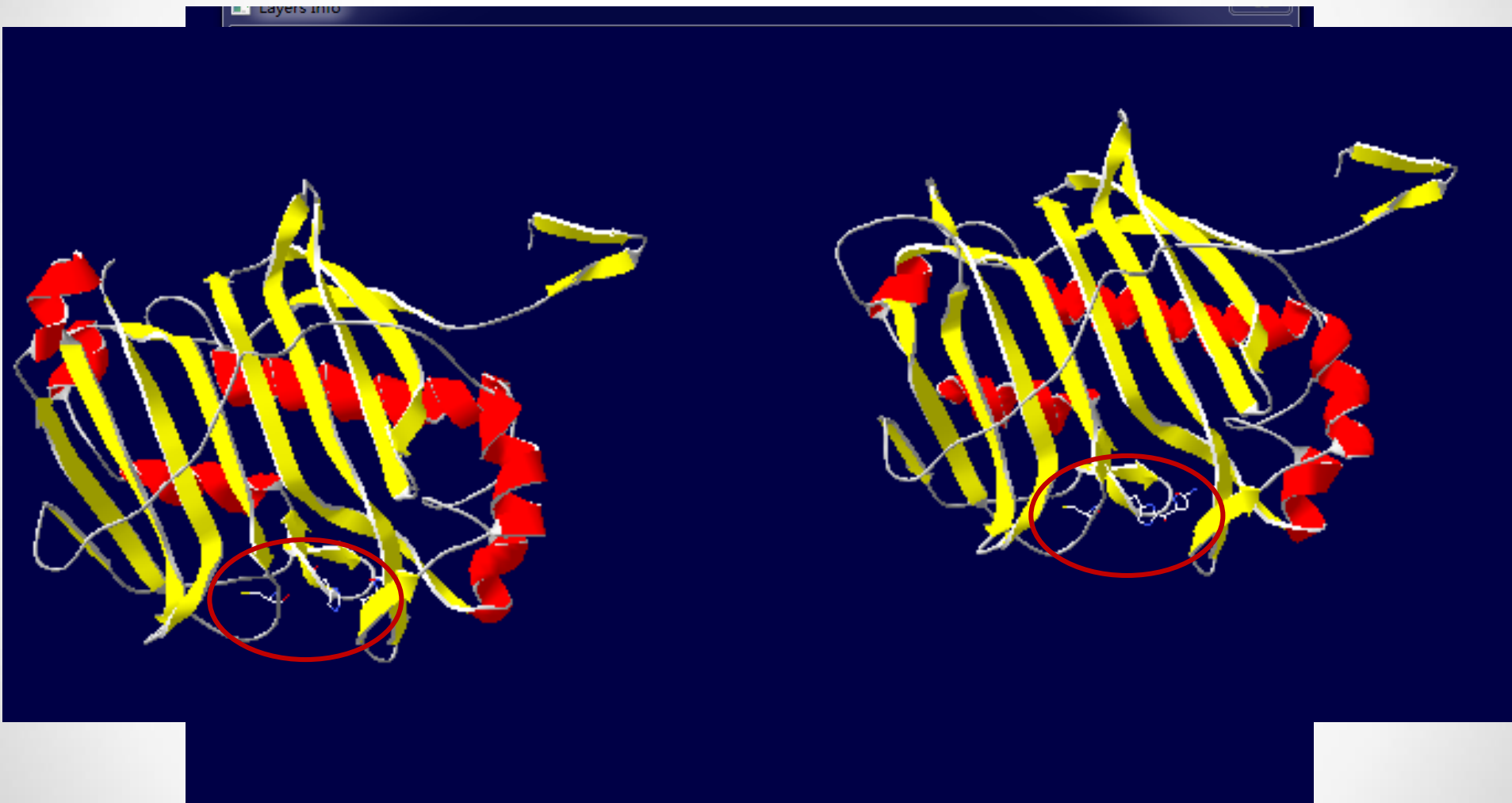


黑框：突变后酶失活



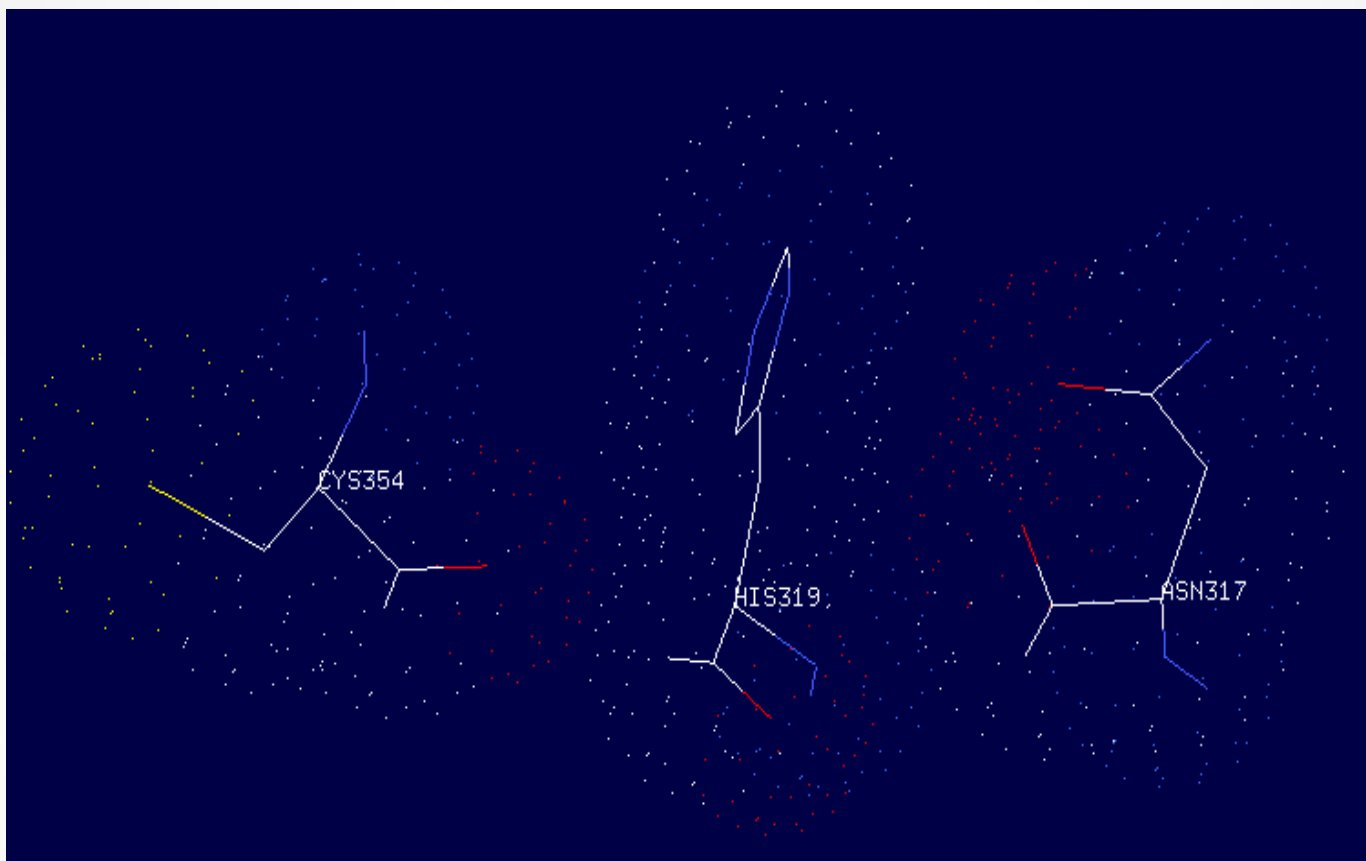
蓝框：影响酶活的关键氨基酸

Cys、Asn、His影响催化活性



棉花

拟南芥



在棉花FatB蛋白序列中Asn317、His319、Cys354为催化活性中心的保守残基。

Unconserved 0 1 2 3 4 5 6 7 8 9 10 Conserved

	10	20	30	40	50	
sp_Q42561_FATA1	-----	-----	-----	-----	ML	
sp_Q9SV64_FATA2	-----	-----	-----	-----	ML	
sp_Q9SJE2_FATB	MVATSATSSF	FPVPSSS	---	LDPNGKGNK	IGSTNLAGLN	SAPNSGRMKV
sp_Q9SQI3_FATB	MVATAVTSAF	FPVTSSPDSS	DSKNKKLG	---	SIKSK	PSVSSGSLQV
tr_A0A1U8M1F9_A	MVATAAATSSF	FPVTSSPDSS	DSKNKKLGSG	STNLGGIKSK	PSASSGSLQV	
tr_A0A288XSA9_A	AMVATAAATSSF	FPITSSPDSI	DSKNKKLNG	STNLGGIKLK	PSASSGSLQV	
Consistency	4444334434	4432442221	222222410	1011112222	2313442368	

	60	70	80	90	100		
sp_Q42561_FATA1	KLSCNVTD SK	LQRSL LFFSH	SYRSD	----	PVNFIRRRIV	SCSQTKKTGL	
sp_Q9SV64_FATA2	KLSCNVTDHI	H----	NLFSN	SRRIF	----	VPVHRQTRPI	SCFQL---KK
sp_Q9SJE2_FATB	KPNAQAPPKI	NGKKVGLPGS	VDIVRTDTET	SSHPAPRTFI	NQLPDWSMLL		
sp_Q9SQI3_FATB	KANAQAPPKI	NGTVASTTPV	EGSKND DG	---	ASSPPPRTFI	NQLPDWSMLL	
tr_A0A1U8M1F9_A	KANAQAPPKI	NGTTVVTSPV	EGFKNE DG	---	AGSPHPRTFI	NQLPDWSMLL	
tr_A0A288XSA9_A	KANAQAPPKI	NGTTVVMTPV	EGFPSE DA	---	ASSLPPRTFI	NQLPDWSMLL	
Consistency	*476676667	5433434343	5333324100	5443257659	7566545447		

	110	120	130	140	150		
sp_Q42561_FATA1	VPLRAVVSAD	QGS	-----	V	VQGLATLADQ	LRLGSLTEDG	LSYKEK FVVR
sp_Q9SV64_FATA2	EPLRAILSAD	HGNSSVRVAD	TVSGTSPADR	LRFGRLMEDG	FSYKEK FIVR		
sp_Q9SJE2_FATB	AAITTIIFLAA	EKQWMMLDWK	PRRSDMLVDP	FGIGRIVQDG	LVFRQNFSIR		
sp_Q9SQI3_FATB	AAITTIIFLAA	EKQWMMLDWK	PRRPDMVIDP	FGIGKIVQDG	LVFSQNF SIR		
tr_A0A1U8M1F9_A	AAITTIIFLAA	EKQWMMLDWK	PRRPDMLIDP	FGIGRIVQDG	LVFRQNFSIR		
tr_A0A288XSA9_A	AAMTTIIFLAA	EKQWMMLDWK	PKRPDMLIDP	FGIGRIVQDG	LVFRQNFSIR		
Consistency	66767966*	5756445445	55535666*	5757*7878*	868686*59*		

	160	170	180	190	200
sp_Q42561_FATA1	SYEVGSNKTA	TVETIANLLQ	EVGCNHAQSV	GFSTDGFATT	TTMRKLHLIWI
sp_Q9SV64_FATA2	SYEVGINKTA	TIETIANLLQ	EVACNHVQNV	GFSTDGFATT	LTMRKLHLIWI
sp_Q9SJE2_FATB	SYEIGADRSA	SIETVMNHLQ	ETALNHVKTA	GLLGDGFGST	PEMFKKNLIWI
sp_Q9SQI3_FATB	SYEIGADQTA	SIETLMNHLQ	ETAINHCRSA	GLLGEFGGAT	PEMCKKNLIWI
tr_A0A1U8M1F9_A	SYEIGADRTA	SIETLMNHLQ	ETAINHCKSA	GLLGEFGGAT	PEMCKKNLIWI
tr_A0A288XSA9_A	SYEIGADRTA	SIETLMNHLQ	ETAINHCKSA	GLLGDGFGAT	PGMCKKNLIWI
Consistency	**9*6778*	79**86*5**	*785**5677	*7658**76*	54*3*57**

162



..... 210. 220. 230. 240. 250

sp_Q42561_FATA1	V	T	A	R	M	H	I	E	I	Y	K	P	A	W	G	D	V	V	E	I	E	T	W	C	Q	S	E	G	R	I	G	T	R	R	D	W	I	L	K	D	S	V	T	G	E	V	T	G	R	
sp_Q9SV64_FATA2	V	T	A	R	M	H	I	E	I	Y	K	P	A	W	S	D	V	V	E	I	E	T	W	C	Q	S	E	G	R	I	G	T	R	R	D	W	I	L	K	D	C	A	T	G	E	V	I	G	R	
sp_Q9SJE2_FATB_	V	V	T	R	M	Q	V	V	V	D	K	Y	P	T	W	G	D	V	V	E	V	D	T	W	V	S	Q	S	G	K	N	G	M	R	R	D	W	L	V	R	D	C	N	T	G	E	T	L	T	R
sp_Q9SQI3_FATB_	V	V	T	R	M	Q	V	V	V	D	R	Y	P	T	W	G	D	V	V	Q	V	D	T	W	V	S	A	S	G	K	N	G	M	R	R	D	W	L	V	S	N	S	E	T	G	E	I	L	T	R
tr_A0A1U8M1F9_A	V	V	T	R	M	Q	V	V	F	D	R	Y	P	T	W	G	D	V	V	Q	V	D	T	W	V	S	A	S	G	K	N	G	M	R	R	D	W	L	V	S	D	S	K	T	G	E	V	L	T	R
tr_A0A288XSA9_A	V	V	T	R	M	Q	V	V	D	C	Y	P	T	W	G	D	V	V	Q	V	D	T	W	V	S	A	S	G	K	N	G	M	R	R	D	W	L	V	S	N	S	K	T	G	E	I	L	T	R	
Consistency	*	7	7	*	*	6	9	5	7	5	5	*	*	7	*	8	*	*	8	9	8	*	*	6	7	6	7	*	8	5	*	6	*	*	*	8	8	6	7	6	3	*	*	*	7	7	5	*		

..... 260. 270. 280. 290. 300

sp_Q42561_FATA1	A	T	S	K	W	V	M	M	N	Q	D	T	R	R	L	Q	K	V	S	D	D	V	R	D	E	Y	L	V	F	C	P	Q	E	P	R	L	A	F	P	E	E	N	N	R	S	L	K	K	I	P
sp_Q9SV64_FATA2	A	T	S	K	W	V	M	M	N	Q	D	T	R	R	L	Q	R	V	T	D	E	V	R	D	E	Y	L	V	F	C	P	P	E	P	R	L	A	F	P	E	E	N	N	S	S	L	K	K	I	P
sp_Q9SJE2_FATB_	A	S	S	V	W	V	M	M	N	K	L	T	R	R	L	S	K	I	P	E	E	V	R	G	E	I	E	P	Y	F	V	N	S	--	D	P	V	L	A	E	D	S	R	K	L	T	K	I	D	
sp_Q9SQI3_FATB_	A	T	S	V	W	V	M	M	N	K	L	T	R	R	L	S	K	I	P	E	E	V	R	G	E	I	E	P	F	F	M	N	S	--	D	P	V	L	A	E	D	S	Q	K	L	V	K	L	D	
tr_A0A1U8M1F9_A	A	S	S	V	W	V	M	M	N	K	L	T	R	R	L	S	K	I	P	E	E	V	R	G	E	I	E	P	Y	F	M	N	S	--	D	P	V	V	A	E	D	S	R	K	L	V	K	L	D	
tr_A0A288XSA9_A	A	S	S	V	W	V	M	M	N	K	L	T	R	R	L	S	K	I	P	E	E	V	R	G	E	I	E	P	H	F	M	N	S	--	D	P	V	V	A	E	D	N	R	K	L	V	K	L	D	
Consistency	*	7	*	5	*	*	*	*	7	4	*	*	*	7	8	9	6	8	8	*	6	*	6	5	5	6	5	4	5	7	0	0	4	6	6	4	6	*	7	7	6	7	*	4	*	8	6			

..... 310. 320. 330. 340. 350

sp_Q42561_FATA1	K	L	E	D	P	A	Q	Y	S	M	I	G	L	K	P	R	R	A	D	L	D	M	N	Q	H	V	N	N	V	T	Y	I	G	W	V	L	E	S	I	P	Q	E	I	V	D	T	H	E	L	Q
sp_Q9SV64_FATA2	K	L	E	D	P	A	Q	Y	S	M	L	G	L	K	P	R	R	A	D	L	D	M	N	Q	H	V	N	N	V	T	Y	I	G	W	V	L	E	S	I	P	Q	E	I	I	D	T	H	E	L	K
sp_Q9SJE2_FATB_	--	D	K	T	A	D	Y	V	R	S	G	L	T	P	R	W	S	D	L	D	V	N	Q	H	V	N	N	V	K	Y	I	G	W	I	L	E	S	A	P	V	G	I	M	E	R	Q	K	L	K	
sp_Q9SQI3_FATB_	--	D	S	T	A	E	H	V	C	K	G	L	T	P	K	W	S	D	L	D	V	N	Q	H	V	N	N	V	K	Y	I	G	W	I	L	E	S	A	P	L	P	I	L	E	S	H	E	L	S	
tr_A0A1U8M1F9_A	--	K	S	M	A	E	H	V	R	K	G	L	T	P	R	W	S	D	L	D	V	N	Q	H	V	N	N	V	K	Y	I	G	W	I	L	E	S	A	P	L	P	V	L	E	T	H	E	L	S	
tr_A0A288XSA9_A	--	D	S	T	A	Q	Y	V	R	K	G	L	T	P	R	W	S	D	L	D	V	N	Q	H	V	N	N	V	K	Y	V	G	W	I	L	E	S	T	P	L	G	I	V	E	S	H	E	L	C	
Consistency	0	0	6	5	4	*	7	7	6	4	4	*	*	6	*	8	5	3	*	*	*7	*	*	*	*	*	*	*	6	*9	*	*	9	*	*	5	*	5	3	9	7	8	6	7	8	*4				

	360	370	380	390	400
sp_Q42561_FATA1	VITLDYRREC	OODDVVDSL	TT-----	TS EIGGTNGSAT	SGTOGHND
sp_Q9SV64_FATA2	VITLDYRREC	QQDDIVDSL	TSETPNEVVS	KLTGTNGSTT	SSKREHNESH
sp_Q9SJE2_FATB	SMTLEYRREC	GRDSVLQSL	AV-----	-----TGCDI	GNLATAGDVE
sp_Q9SQI3_FATB	ALTLEYRREC	GRDSVLQSL	TV-----	-----SDSNT	ENAVNVGEFN
tr_A0A1U8M1F9_A	SMTLEYRREC	GRESILQSL	TV-----	-----SDSSV	GDLVNVGEIE
tr_A0A288XSA9_A	SMTLEYRREC	GRDSVLQSL	AV-----	-----SGVGN	LGNM--GEIE
Consistency	57**8****	5787986**	7600000000	0000066545	3534226845

	410	420	430	440
sp_Q42561_FATA1	FLHLLRLSGD	GQEINRGTTL	WRKKPSS	-----
sp_Q9SV64_FATA2	FLHILRLSEN	GQEINRGRTQ	WRKKSSR	-----
sp_Q9SJE2_FATB	CQHLLRLQ-D	GAEVVRGRTE	WSSKTPT	TTW GTAP-----
sp_Q9SQI3_FATB	CQHLLRLD-D	GAEIVRGRTR	WRPKHAK	SSA NMDQITAKRA
tr_A0A1U8M1F9_A	CQHLLQLE-E	GSEIVRGRTQ	WRPKYAK	SFG NVGQIPAESA
tr_A0A288XSA9_A	CQHLLQLE-E	GSEIVRGRTQ	WRPKNAK	SFG KMDQVPAQSA
Consistency	55*9*7*506	*5*95**7*5	*75*365311	1212112112

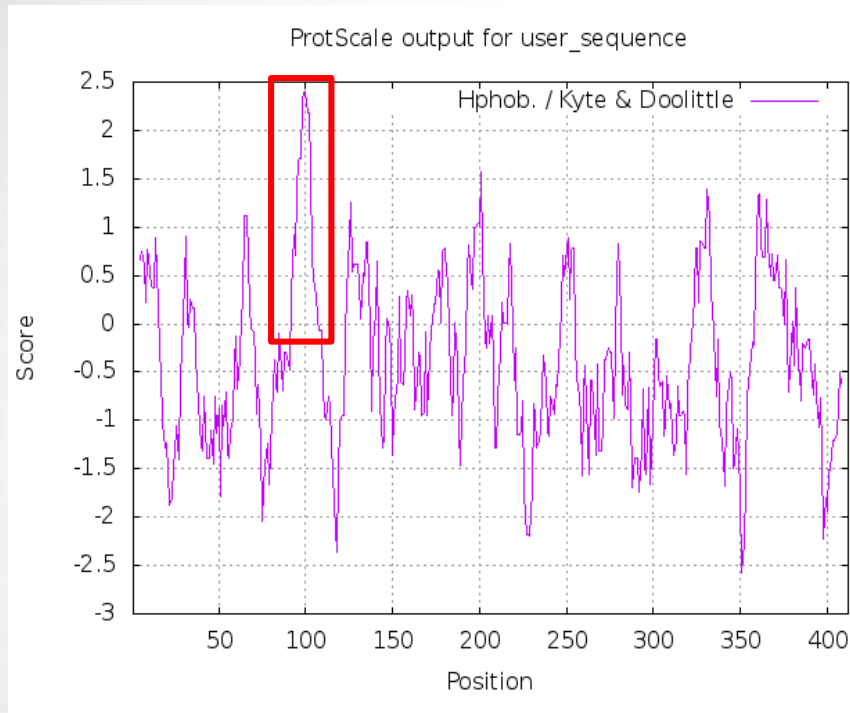
	162	201	262	317
AtFatB	Ile	Val	Ser	Trp
GhFatB	Ile	Val	Ser	Trp
AtFatA1	Val	Thr	Gln	Arg
AtFatA2	Ile	Thr	Gln	Arg

AtFatBh和GhFatB与底物特异性有关的4个残基相同，它们可能具有相同的底物特异性。

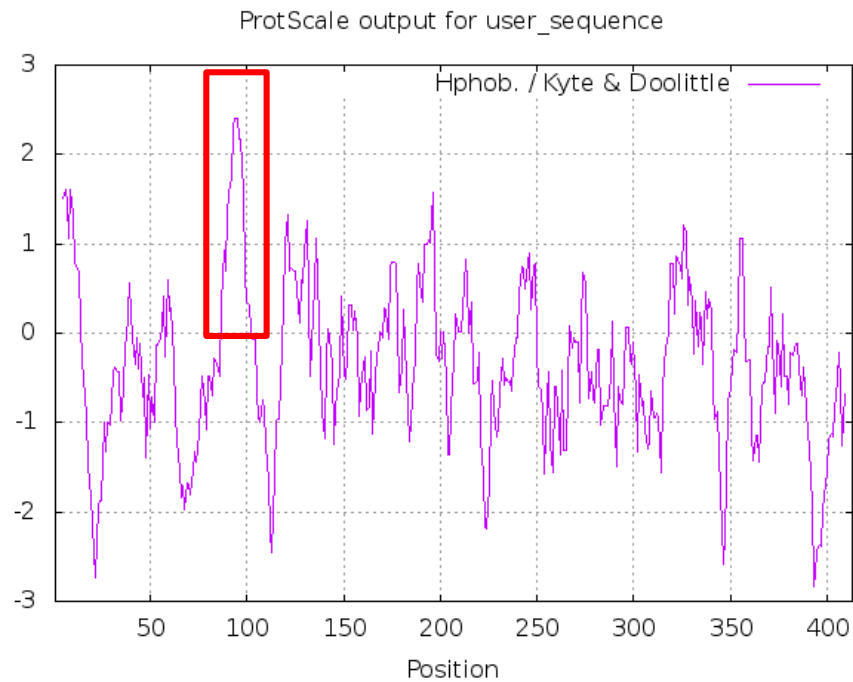
FatA和FatB的4个残基比较，也说明这4个残基位点可能与Fat蛋白的底物特异性有关。

棉花FatB蛋白结构分析

疏水性预测



AtFatB

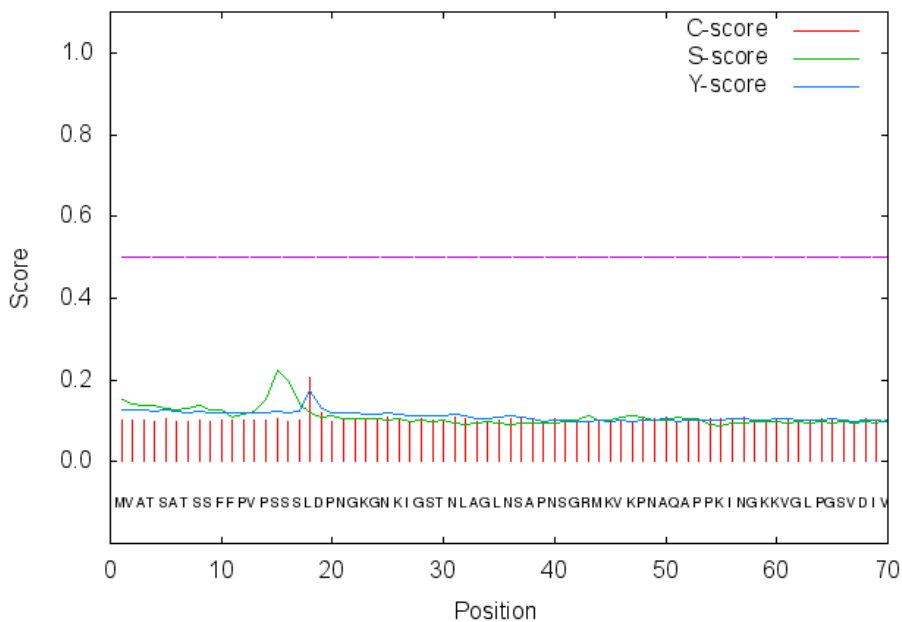


GhFatB

与Facciotti等研究结果一致

信号肽预测

SignalP-4.1 prediction (euk networks): sp_Q9SJE2_Arabidopsis

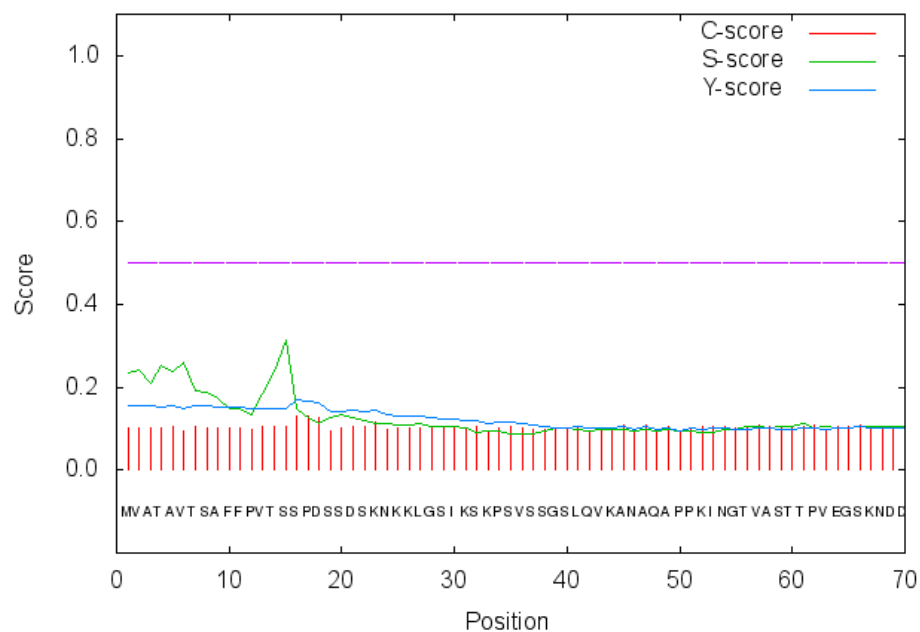


# Measure	Position	Value	Cutoff	signal peptide?
max. C	18	0.206		
max. Y	18	0.172		
max. S	15	0.223		
mean S	1-17	0.141		
D	1-17	0.155	0.450	NO

Name=sp_Q9SJE2_Arabidopsis SP='NO' D=0.155 D-cut off=0.450 Networks=SignalP-noTM

AtFatB

SignalP-4.1 prediction (euk networks): sp_Q9SQI3_Gossypium



# Measure	Position	Value	Cutoff	signal peptide?
max. C	16	0.132		
max. Y	16	0.170		
max. S	15	0.312		
mean S	1-15	0.210		
D	1-15	0.192	0.450	NO

Name=sp_Q9SQI3_Gossypium SF='NO' D=0.192 D-cut off=0.450 Networks=SignalP-noTM

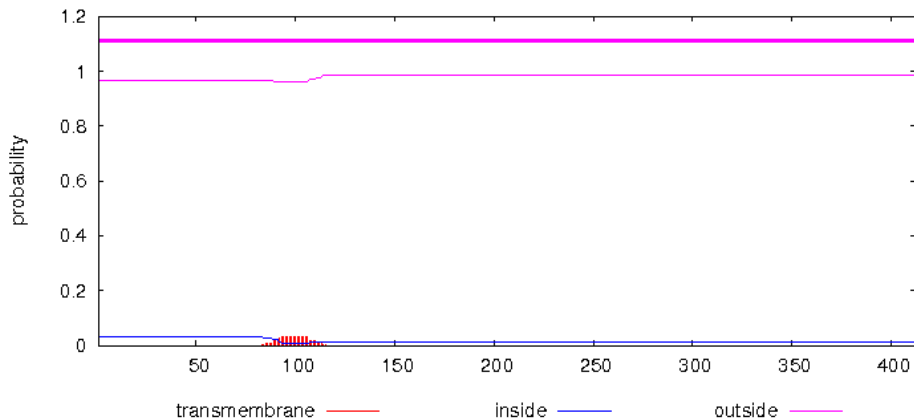
GhFatB

跨膜螺旋预测

```
# sp|Q9SJE2|Arabidopsis Length: 412
# sp|Q9SJE2|Arabidopsis Number of predicted TMHs: 0
# sp|Q9SJE2|Arabidopsis Exp number of AAs in TMHs: 0.664800000000000001
# sp|Q9SJE2|Arabidopsis Exp number, first 60 AAs: 0
# sp|Q9SJE2|Arabidopsis Total prob of N-in: 0.03134
sp|Q9SJE2|Arabidopsis TMHMM2.0 outside 1 412
```

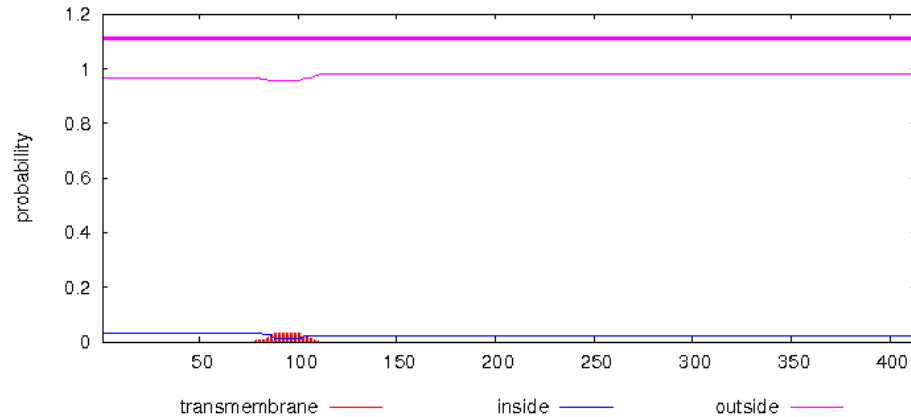
```
# sp|Q9SQI3|Gossypium Length: 413
# sp|Q9SQI3|Gossypium Number of predicted TMHs: 0
# sp|Q9SQI3|Gossypium Exp number of AAs in TMHs: 0.665360000000000001
# sp|Q9SQI3|Gossypium Exp number, first 60 AAs: 0.00018
# sp|Q9SQI3|Gossypium Total prob of N-in: 0.03181
sp|Q9SQI3|Gossypium TMHMM2.0 outside 1 413
```

TMHMM posterior probabilities for sp|Q9SJE2|Arabidopsis



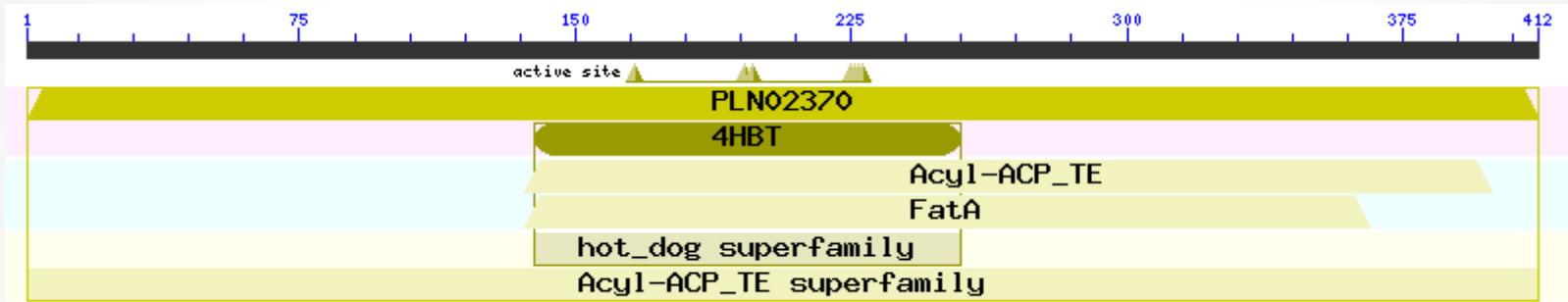
AtFatB

TMHMM posterior probabilities for sp|Q9SQI3|Gossypium

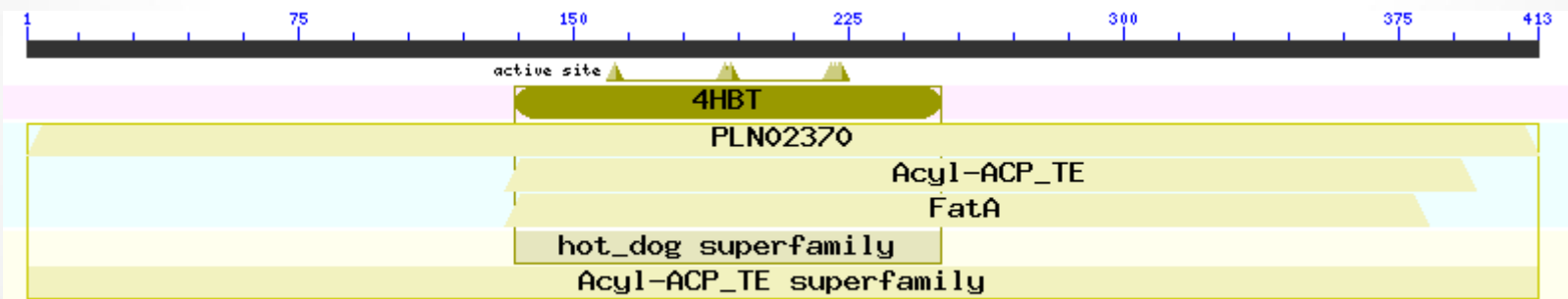


GhFatB

结构域预测

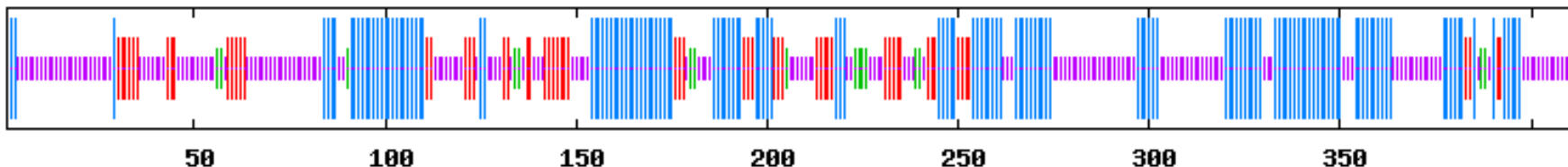
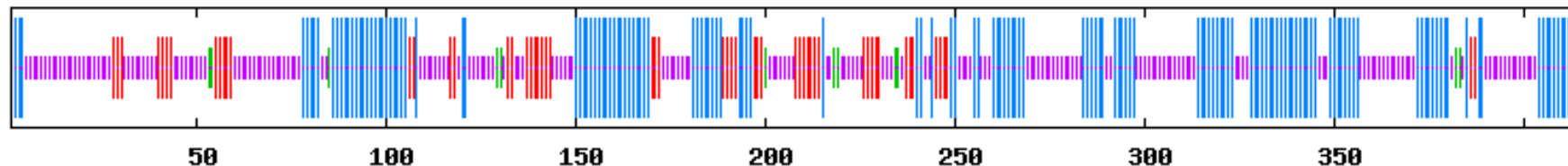


拟南芥FatB1编码蛋白质的结构域预测



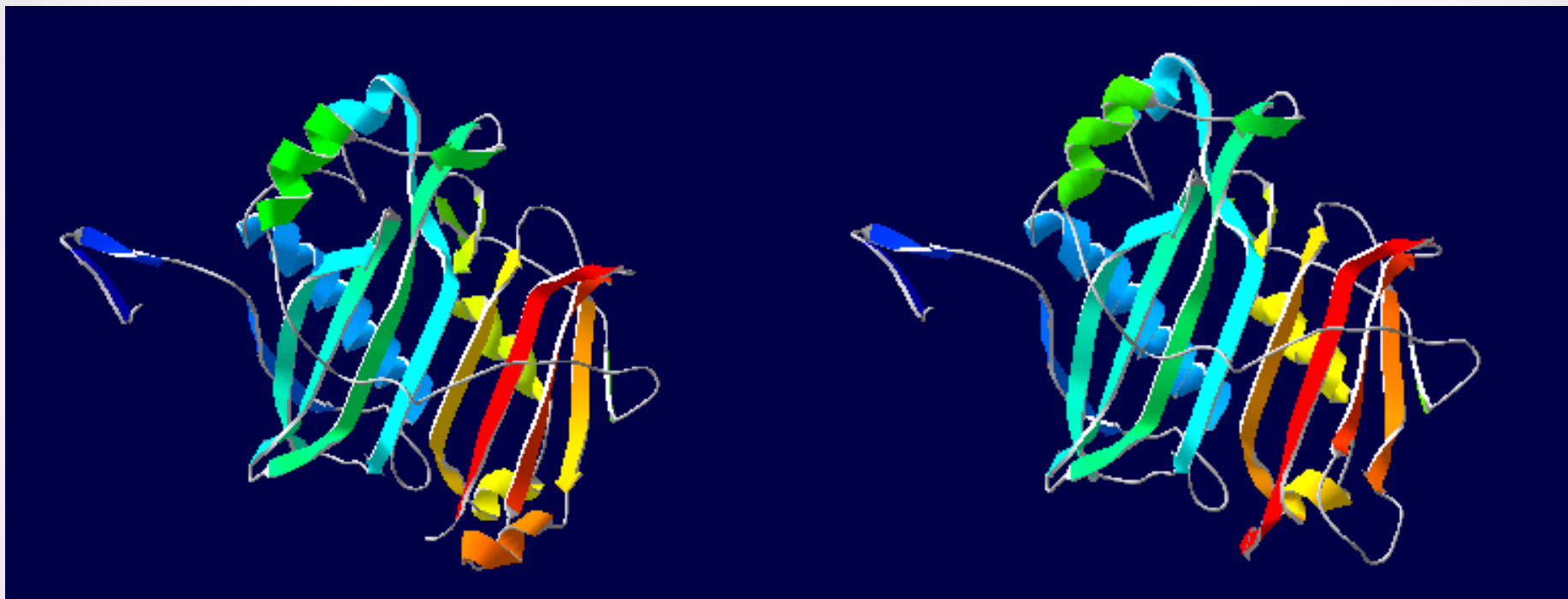
陆地棉FatB1编码蛋白质的结构域预测

二级结构



	AtFatB	GhFatB
α 螺旋	36.80%	35.68%
延伸链	14.04%	15.05%
β 转角	2.91%	3.88%
无规卷曲	46.97%	45.39%

三级结构



AtFatB

GhFatB

两个热狗状结构域（1个螺旋/4个折叠片层）

结论

- 棉花FatB蛋白是一个非转运蛋白，无跨膜结构。
- 棉花FatB蛋白与拟南芥FatB蛋白高度相似，催化活性中心的关键氨基酸一致，具有FatB蛋白的功能作用。
- 预测了棉花FatB蛋白中影响底物特异性的保守氨基酸残基。
 -

待解决问题

- 通过蛋白结构预测，我们推测棉花FatB蛋白与拟南芥FatB蛋白可能功能相似。但棉花是异源四倍体植物，其遗传结构更为复杂，需要通过详细的试验验证。

参考文献

- Mayer KM. A structural model of the plant acyl-acyl carrier protein thioesterase FatB comprises two helix/4-stranded sheet domains, the N-terminal domain containing residues that affect specificity and the C-terminal domain containing catalytic residues[J]. *Journal of Biological Chemistry*, 2005, 280(5):3621-3627.
- Bonaventure. Disruption of the FATB Gene in Arabidopsis Demonstrates an Essential Role of Saturated Fatty Acids in Plant Growth[J]. *Plant Cell*, 2003, 15(4):1020-1033.
- 王威浩. 油茶FatB基因的全长cDNA克隆[D]. 中南林业科技大学, 2009.

感谢罗老师的教导
和各位小组成员的共同努力！